

Development of Prognosis Factors in a Scoring System for Predicting of Breast Cancer Mortality

Furat N. Tawfeeq

Media & Informatics Unit, National Cancer Research Center, University of Baghdad, Baghdad, Iraq

E-mail: furatnidhal@yahoo.com

Abstract

Today, the prediction system and survival rate became an important request. A previous paper constructed a scoring system to predict breast cancer mortality at 5 to 10 years by using age, personal history of breast cancer, grade, TNM stage and multicentricity as prognostic factors in Spain population. This paper highlights the improvement of survival prediction by using fuzzy logic, through upgrading the scoring system to make it more accurate and efficient in cases of unknown factors, age groups, and in the way of how to calculate the final score. By using Matlab as a simulator, the result shows a wide variation in the possibility of values for calculating the risk percentage instead of only 16. Additionally, the accuracy will be calculated with risk mortality at 5 and 10 years depending on the number of unknown factors. The new system presented in a graphical user interface to facilitate the friendliness at the user side.

Keywords: Risk calculation, scoring system, fuzzy set, breast cancer, prognostic factor, risk prediction

1. Introduction

In developed and developing regions of the world, Breast cancer still the most diagnosed tumour and have the main cause of cancer death in women (Globocan 2012). Much information about the attitude of breast cancer (demographic, clinical, pathological and molecular) could be provided by a large number of prognostic factors which empower the relation between the findings and treatments. Among them, lymph node (N), tumour size (T) and grade are the most critical factors (Soerjomataram *et al.* 2008; Fisher *et al.* 2001; Williams *et al.* 2006).

Survivability is the ability to stay alive, the statistics that deals with the modelling of time to event data are called survival analysis; in such case, an event is a death. Tumour characteristics can give a hint in calculating survival time. The survival rate refers to the percentage of patients who live at least 5 years after being diagnosed with cancer (Nivetha & Samundeeswari 2016).

Nowadays, clinical decision-making plays a vital role in finding the risk and survival duration for patients suffering from malignant disease (Chen *et al.* 2009), by selecting accurate prognostic factors, an artificial intelligence could be used to calculate the risk and then inform the patient.

Zhenchong X. *et al.* (2018) built a nomogram model to predict one to three years survival probability of the patients with metastatic breast cancer (MBC) depending on the results of the multivariable analysis with the Cox hazard model, the survival probability supposed to be gain from the summation of factor's scores. Esther *et al.* (2017) constructed a system to calculate scoring points, where each prognosis factor has weight, the final score will predict breast cancer mortality at 5 to 10 years. Azadeh *et al.* (2017) used machine learning in improving the survival prediction based on gene expression, by collecting articles from 2000 to 2016 on survival, cancer and machine learning. Nivetha & Samundeeswari (2016) made a comparison between support vector machine SVM and fuzzy rule-based in the predicting survival of breast cancer patients; they found that fuzzy is better than SVM in the classification.

Esther *et al.* (2017) produced a scoring system for predicting breast cancer mortality at 5 to 10 years, by collecting data from patients diagnosed with breast cancer in the Elda Health Department between 2003 to 2006, the sample size was 287 women, of whom 55 died of breast cancer in the Spanish region, the main parameter was time-to-death, a point scoring system was used to build a model for calculating and predicting the survival rate by using the Framingham Heart Study methodology. This system selects the prognostic factors that effect on the mortality and risk calculation, then, giving them a score weight. The sum of these scores or points produces the risk of death (Sullivan *et al.* 2004). By this, the model covert the real information collected from patients to a predictive scoring system.

Esther *et al.* (2017) found that age, history of breast cancer, grade, stage, and multicentricity are the main factors that increase the risk mortality through his and previous studies (Soerjomataram *et al.* 2008; Jimenez *et al.* 2003;

Galea *et al.* 1992; Bloom & Richardson 1957; Frederick *et al.* 2010; Louwman *et al.* 2005; Vera-Badillo *et al.* 2014). Fig. 1 shows the model of a scoring system to predict breast cancer mortality.

Age (years) Points		Stage Points		Risk (%)		
<50	0	0	0	Point sum	5 years	10 years
≥50	1	I	1	1	0.38	0.68
Personal history of breast cancer Points		II	3	2	0.69	1.24
No	0	III	5	3	1.26	2.25
Yes	3	IV	7	4	2.30	4.08
Grade Points		Multicentricity Points		5	4.16	7.34
Low	0	No	0	6	7.48	13.02
Intermediate/moderate	1	Yes	1	7	13.26	22.53
High	3			8	22.92	37.31
				9	37.89	57.45
				10	58.15	79.05
				11	79.68	94.27
				12	94.58	99.47
				13	99.52	99.99
				14	99.99	>99.99
				15	>99.99	>99.99

Figure 1: Scoring system to predict breast cancer mortality (Esther *et al.* 2017)

From Fig. 1, we can notice that the age score was divided into two groups; above or equal 50 and less than 50 years, giving them one and zero points respectively, which are very crisp value, for example, if patient was 49 years old, then the point for age group will be zero, while it was very close to 50, by other words the patients with age near 50 should be affected by the values of these two groups. Additionally, the system works only and only if the five predictive factors were known and when anyone was missing then the system will give us false risk percentage.

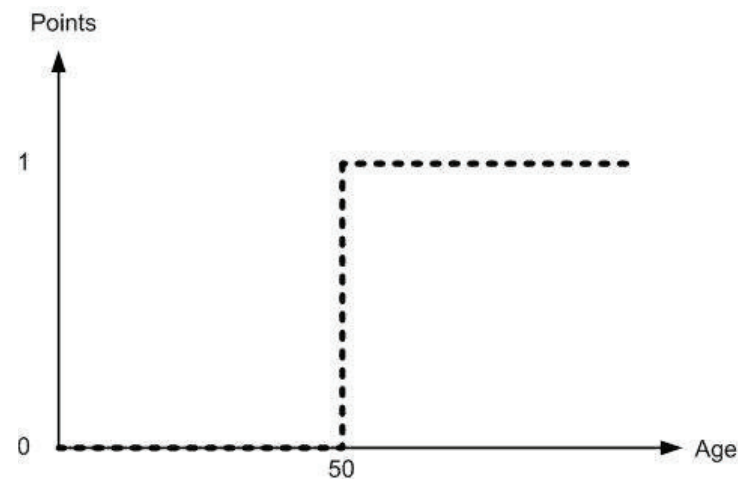
This paper tries to improve this scoring system by the embedded fuzzy controller in the process of finding risk calculation to reduce the limitation mention before and to make the model more sophisticated.

Many techniques were used recently in prediction models, especially in the cancer field, to improve clinical decision making and to convert laboratory finding into information that helps in informing the patient and in caregiving by the medical staff.

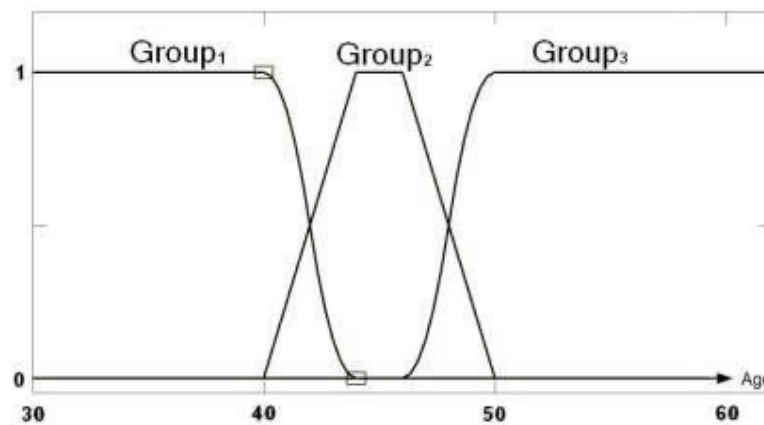
The fuzzy controller can give an accurate classification through a set of rules in-form of IF-THEN instead of using crisp sets as in classical rules. Fuzzy sets have degrees of membership functions, which have values between one and zero. A degree of zero means that an object is not a member of that set, while one, means that it is, the most important part which distinguishes fuzzy sets are the values that lay in-between one and zero, that will have a partial degree of membership (Bai *et al.* 2006).

2. Materials and Methods

This paper improves the point system mentioned in (Esther *et al.* 2017) through three steps, the first step deals with the age group. As mention before, (Esther *et al.* 2017) divided the age into two groups as shown in Fig. 1, while in this system, each crisp value will be mapped into multiple fuzzy variables, and this process is called fuzzification. Fig. 2A shows the how to find the value for age factor according to (Esther *et al.* 2017) point system, while Fig. 2B describes the transformation of crisp age into fuzzy sets.



(A)



(B)

Figure 2: Convert crisp age value (A) into age fuzzy set (B).

By this, the output point could be any value between zero and one instead of crisp values, the age between 40 and 50 will belong to two fuzzy sets due to overlapping. Fig. 3A, 3B and 3C show how to calculate the fuzzy membership values when the fuzzy sets are trapezoidal, R and L-function respectively, which are used in this paper.

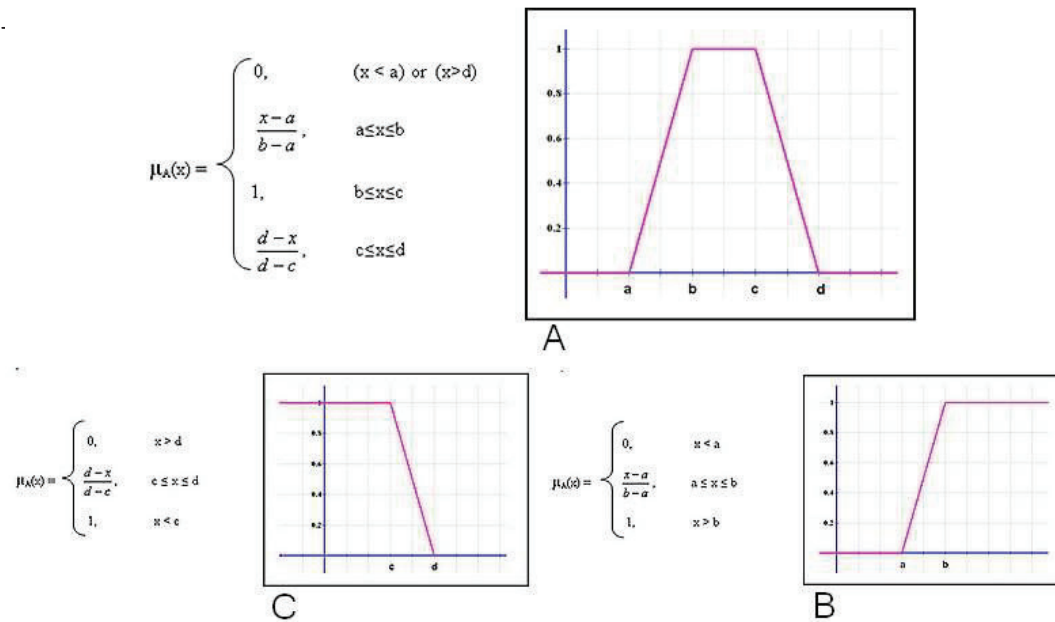


Figure 3: How to calculate fuzzy membership value when the fuzzy set is (A) trapezoidal, (B) R-function, (C) L-function (eMathTeacher 2018)

Where a, d, b, and c are the lower limit, upper limit, lower support limit, and upper support limit respectively. To determine the output, rules should be chosen in a proper way to verify the results, the rules, in this case, are shown in Table 1.

Table 1: Rules for age point

No.	Rules
R1	IF Age is less than or equal 40 THEN point is zero
R2	IF Age is greater than or equal 50 THEN point is one
R3	IF Age is equal to 45 THEN point is 0.5

From above, the output of age point will be a value between and including zero or one, while in (Esther *et al.* 2017); the point is zero or one. The resultant fraction point will be solved later before getting the mortality risk percentage.

3. Results and Discussion

As mentioned before, this paper enhanced the prediction in three steps; the second one is entering new point for the unknown factors (Age, personal history of breast cancer, grade, stage, and multicentricity). These modifications take place for any missing information which without it, the whole scoring system will be crashed. Fig. 4 defines these modifications.

Age (years)	Points
Unknown	0.5

Personal history of breast cancer	Points
No	0
Yes	3
Unknown	1.5

Grade	Points
Low	0
Intermediate/moderate	1
High	3
Unknown	1.5

Stage	Points
0	0
I	1
II	3
III	5
IV	7
Unknown	3.5

Multicentricity	Points
No	0
Yes	1
Unknown	0.5

Figure 4: Adding unknown choices to the scoring system

The unknown point was selected to be in the middle of upper and lower limit, for example, the limits for the grade are 0 and 3, so the unknown point will be 1.5.

After sum all the points, the result will be number between zero and 15, with the possibility of a fraction. Esther *et al.* (2017) used the scoring system in Fig. 1 to calculate risk percentage at 5 to 10 years. But now, after adding unknown choices and change the way to find age point, modifications on the risk finding should take place, which represents the third step in improvement. Again, fuzzy sets will benefit from the existing system in defining the rules. The rules were very similar to what was exists in Fig. 1, but the result will be different, because of fraction point. Table 2 shows rules 5, 6 and 7 for finding the percentage of risk mortality.

Table 2: Some of the rules for risk (%)

No.	Rules
R5	IF point sum= 4 THEN 5 years risk=4.16 and 10 years risk=7.34
R6	IF point sum= 5 THEN 5 years risk=7.48 and 10 years risk=13.02
R7	IF point sum= 6 THEN 5 years risk=13.26 and 10 years risk=22.53

The input of the fuzzy interface system (FIS) has 15 membership functions with a range from zero to 15 that represent the main point sum, while the output of FIS consists of 15 membership functions with a range from zero to 100, that represent the values of risk (%) mentioned in Fig. 1.

The accuracy of this system could be measured depends on the number of unknown factors, the relation between them is inverse; when the unknown factor increases, then the accuracy decrease. The equation to find the accuracy is as shown in Equation 1.

$$Accuracy = \begin{cases} 100 & \text{for } N=0 \\ 100-(20 \cdot N) & \text{for } N \neq 0 \end{cases} \quad (1)$$

Where N represents the number of the unknown factors. To make the system more friendliness, a graphical user interface (GUI) was used as shown in Fig. 5, also the unknown choice was added to the prognostic factors. As a result, the final outputs are point sum, risk (%) at 5 to 10 years and the accuracy (%) of the system.

A scoring system to predict breast cancer mortality at 5 to 10 years

Age Type Un for unknown
 Fraction could be used

Do you have history of breast cancer

Grade

Stage

Multicentricity

Point sum

Risk (%) at 5 years

Risk (%) at 10 years

Accuracy (%)

Figure 5: Graphical user interface for our system

Fig. 6 illustrates the flow chart used in (Esther *et al.* 2017), while Fig. 7 shows the revised version proposed in this paper, and by these figures, we can highlight the differences between them, certainly, they both try to predict the mortality at 5 and 10 years.

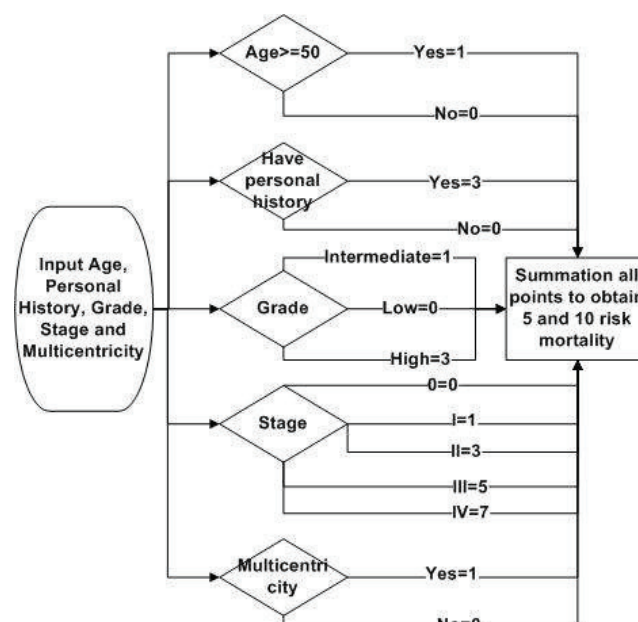


Figure 6: Flow chart of the system used in (Esther *et al.* 2017)

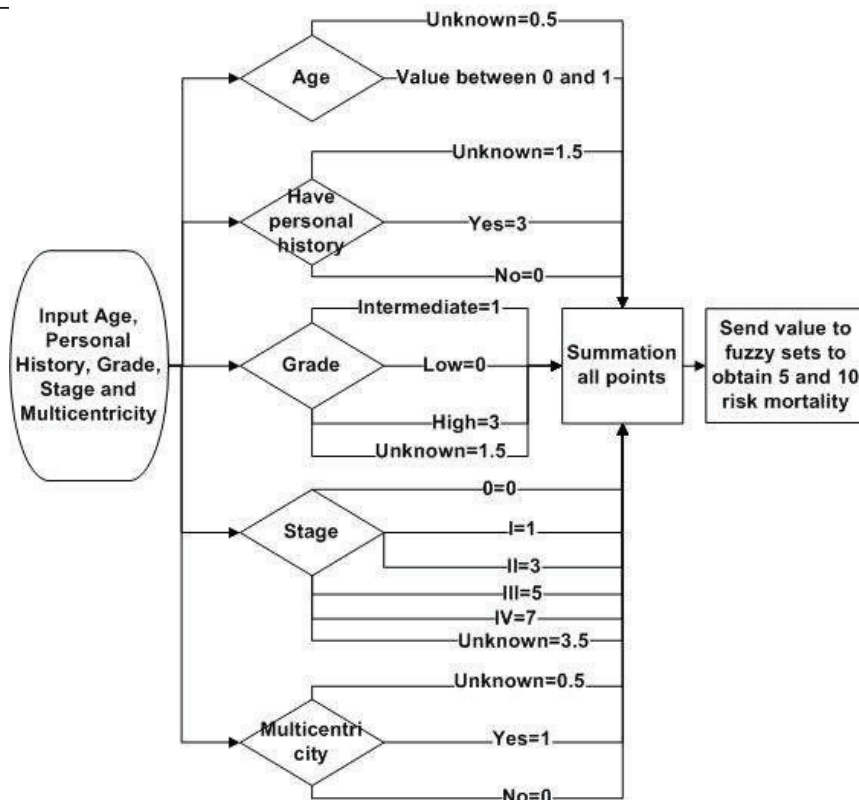


Figure 7: Flow chart of the proposed system

4. Conclusions

An enhancement of the point system for predicting mortality due to breast cancer at 5 to 10 years has been built. The modified system further divides the risk levels to increase the subdivisions, this leads to more variety in risk values; where the original system has only 15 values for risk mortality, while in the new system presented in this paper, there is an infinite number of values due to using of the fuzzy controller. An unknown choice was added to the prognostic factors to make the point system more accurate in case of missing information.

5. Future Work

As mentioned previously, this pointing system works on Spain population, we encourage researchers to find their prognostic factors related to their follow-up patients information because it differs from region to region. Of course, by increasing the number of factors, we will get more accurate results in predicting mortality at 5 and 10 years.

6. Conflict of Interest

The authors declare that they have no conflicts of interest.

References

- Azadeh B., Marjan G., Reza S., Leila S., and Hamide E. (2017), "Improving the Prediction of Survival in Cancer Patients by Using Machine Learning Techniques: Experience of Gene Expression Data: A Narrative Review", *Iran J Public Health*, 46(2), 165-172.
- Bai Y., Zhuang H., Wang, D. (Eds.), (2006), *Advanced Fuzzy Logic Technologies in Industrial Applications*, Chapter 2, Springer, 22-28.
- Bloom HJ. and Richardson WW. (1957), "Histological grading and prognosis in breast cancer; a study of 1409 cases of which 359 have been followed for 15 years", *Br. J. Cancer*, 11, 359-377.
- Chen Y-C., Yang W-W., Chiu H-W. (2009), "Artificial neural network prediction for cancer survival time by gene expression data", in *3rd International Conference on Bioinformatics and Biomedical Engineering*, IEEE.
- eMathTeacher: Mamdani's fuzzy inference method [Online]. (2018), Available: http://www.dma.fi.upm.es/java/fuzzy/fuzzyinf/funpert_en.htm

- Esther P., Antonio P., David M., José R., Antonio F., and Vicente F., (2017), "A scoring system to predict breast cancer mortality at 5 and 10 years", *Scientific Reports* | 7: 415 |DOI:10.1038/s41598-017-00536-7, www.nature.com/scientificreports.
- Fisher ER., Anderson S., Tan-Chiu E., Fisher B., Eaton L. and Wolmark N. (2001), "Fifteen-year prognostic discriminants for invasive breast carcinoma: National Surgical Adjuvant Breast and Bowel Project Protocol-06", *Cancer*, 91 (8 Suppl), 1679–1687.
- Frederick L., David L., Irvin D., April G., Charles M., Daniel G. and Monica M. (2010), *AJCC Cancer Staging Manual*, 2nd ed, Springer, New York.
- Galea MH., Blamey RW., Elston CE. and Ellis IO. (1992), "The Nottingham Prognostic Index in primary breast cancer", *Breast Cancer Res. Treat.*, 22(3), 207–219.
- Globocan 2012, International Agency for Research on Cancer, Lyon, IARC Press, 2013.
- Jimenez-Lee R., Ham B., Vetto J. and Pommier R.. (2003), "Breast cancer severity score is an innovative system for prognosis", *Am. J. Surg*, 186, 404-408.
- Louwman WJ., Janssen H., ML, Houterman, S, Voogd, AC, Van der Sangen, MJC, Nieuwenhuijzen, GAP and Coebergh, JWW. (2005), "Less extensive treatment and inferior prognosis for breast cancer patient with comorbidity: a population-based study", *Eur. J. Cancer*, 41(5), 779–785.
- Nivetha S., Samundeeswari E.S. (2016), "PREDICTING SURVIVAL OF BREAST CANCER PATIENTS USING FUZZY RULE BASED SYSTEM", *International Research Journal of Engineering and Technology*, 3(3), 962-969.
- Soerjomataram I., Louwman M. W., Ribot J. G., Roukema J. A. and Coebergh JW. (2008), "An overview of prognostic factors for long-term survivors of breast cancer", *Breast Cancer Res. Treat.* 107, 309–330.
- Sullivan LM., Massaro JM. and D'Agostino RB. Sr. (2004), "Presentation of multivariate data for clinical use: The Framingham Study risk score functions", *Stat. Med.*, 23, 1631–1660.
- Vera-Badillo FE., Napoleone M., Ocana A., Templeton AJ., Seruga B., Al-Mubarak M., AlHashem H., Tannock IF., and Amir E. (2014), "Effect of multifocality and multicentricity on outcome in early stage breast cancer: a systematic review and meta-analysis", *Breast Cancer Res. Treat.*, 146(2), 235–244.
- Williams C., Brunskill S., Altman D., Briggs A., Campbell H., Clarke M., Glanville J., Gray A., Harris A., Johnston K., and Lodge M. (2006), "Cost-effectiveness of using prognostic information to select women with breast cancer for adjuvant systemic therapy", *Health Technol Assess*, 10(34): iii–iv, ix–xi, 1–204.
- Zhenchong X., Guangzheng D., Xinjian H., Xing L., Xinhua X., Jin W., Zeyu S., and Xi W. (2018), "Score for the Survival Probability in Metastasis Breast Cancer: A Nomogram-Based Risk Assessment Model", *CANCER RESEARCH AND TREATMENT*, [Epub ahead of print].